The C-test as predictor of the academic success of international students

Article

Accepted Version

Michael Daller[1], Amanda Müller[2] and Yixin Wang-Taylor[3]

1 University of Reading (corresponding author)

2 Flinders University

3 Nankei University

The C-test as predictor of the academic success of international students

Abstract

The present article gives an overview of several studies on the predictive validity of the C-test. In the first part of the article, we discuss the construct validity of this test format. Only if the underlying construct of this test is understood, can a justification for high predictive validity be made. In the second part, we discuss several previous studies where the C-test format is used to predict the study and training success of international students. The third part discusses the findings of two as yet unpublished studies on the predictive validity of the C-test. We wish to contribute to the ongoing discussion of the validity of the C-test and argue that it is not only a language test, but also a test of processing speed which is related to working memory. For international students, processing speed in English as a foreign language is related to vocabulary knowledge, which includes statistical knowledge about the probability of words occurring in a given context as well as the probability of words following or preceding each other. The C-test taps precisely into these aspects of language proficiency which explains its high predictive validity for the study success of international students.

Introduction

The latest figures on international students world-wide show a rise from 2 million in 2000 to 5.3 million in 2017 (UNESCO, 2019). The most popular destinations are the US, the UK, and Australia (Universities UK, International 2019). The use of standardised tests for admission, such as IELTS and TOEFL, is rapidly growing in English-speaking countries, with 3.5

million IELTS tests being taken in 2018 (TakeIELTS, 2019). However, these tests are in many cases only weak predictors of the study success of students (for an overview, see Daller & Phelan, 2013; Daller & Yixin, 2017). One can argue that these tests provide a good cut-off point below which students are at risk of failing their studies, but that they are not meant to predict actual study success (the marks the students get). One reason is that the variability of the test scores is truncated because many commencing university students have roughly the same scores. It is therefore difficult to use these truncated tests scores as predictors purely from a mathematical point of view. Daller and colleagues (Daller & Xue, 2009; Daller & Phelan, 2013; Yixin & Daller, 2015; Muller & Daller, 2019; Wang-Taylor, Y., & Milton, J , 2019) have shown that the format is a good alternative to the established tests for the prediction of study success of international students. In the following section, we give an overview of the C-test format and discuss the validity of the test.

The C-test format

The C-test format was developed by Raatz and Klein-Braley (Raatz & Klein-Braley, 1981; Klein-Braley, 1985)[1] as an alternative to the Cloze test, which is based on the deletion of whole words in a text. Instead of deleting whole words, the C-test format - in its classical form - is based on the deletion of only the second half of every second word. Whereas a Cloze test is often based on only a single text, a C-test normally contains five independent sub-texts with 20 gaps each. The use of five texts allows for a variety of subject content, and therefore, can be used to avoid a text bias towards a specific topic. Exact scoring is recommended as there is usually only one possible solution for each gap. Like the Cloze test, the C-test is based on the concept of "reduced redundancy" (Babaii & Ansary, 2001; Spolsky, 1985). There are many redundant elements in natural language, and a native speaker can, to a large extent, restore distorted parts of language input. In a similar vein, Oller (1976) theorised that there was an "expectancy grammar" that allows the native speaker to decode distorted information because they have an intuitive expectation about the transitional probability

[1] A test format that is quite similar to the C-test was already developed in the 19th century: Ebbinghaus, H. (1897): über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern (a new method to test cognitive abilities and its application with school children - translation M Daller). Zeitschrift für Psychologie und Physiologie der Sinnesorgane (Sonder-Abdruck/ Special issue). We have no information about whether the psychologist Raatz knew about this publication, but Ebbinghaus is certainly one of the canonical authors in the German psychological literature.

between linguistic items. However, an operationalisation of these two concepts is yet to be seen.

In a discussion of the validity of the C-test, Klein-Braley (1997) argued that this test format is based on the principle of reduced redundancy and the concept of "expectancy grammar". It has also been argued that the C-test is a test of "general language proficiency" because it correlates highly with other test scores in the four traditional skills (Eckes & Grotjahn, 2006; Klein-Braley 1985). Alderson (2002: 21) argued that there is no general "unitary competence", whereas others have assumed a single construct because of the often high correlations between grammar and vocabulary tests (Singleton & Singleton, 2002: 154). This position is supported by many studies in which C-test scores correlate highly with various other aspects of language proficiency (Bolten, 1992; Dörnyei & Katona, 1992; Grotjahn & Allner, 1996; Hastings, 2002; Huhta, 1996; Jafarpur, 2002). Nevertheless, the assumption of a single "general language proficiency" has not been operationalised, and therefore, the construct validity of the C-test is still being debated. Thus, the operationalisation of the construct validity of this test format is necessary.

More recent studies on the construct validity of the C-test have attempt to operationalise it in a more detailed and measurable way. For native speakers, Wockenfuß and Raatz (2014) came to the conclusion that C-test performance is dependent on processing speed and verbal intelligence. Baghaei and Tabatabaee (2015) argued that the "C-test closely matches the abilities underlying the language component of crystallized intelligence" (2015: 46), or knowledge of facts based on our previous experience which is closely linked to our in-depth vocabulary knowledge. There are also new approaches to operationalising "expectancy grammar" or related concepts. One approach is that of "predictive processing" (Hopp, 2015, 2016) in which statistical knowledge of frequencies in language allows the prediction of possible future items based on probabilistic rules. L1 speakers and advanced language learners have access to this statistical knowledge. For example, the verbs "disappear" and "vanish" are both intransitive, but "disappear" is more frequent, and therefore, learners are far more certain that it cannot take an object. As a consequence, over-generalisations in a learner's language where these intransitive verbs take an object, are far more frequent for "vanish" than for "disappear" (Boyd & Goldberg, 2011: 56). The general theoretical background for these predictions is based on statistical knowledge formulated in the stochastic model of a Markov chain (Gilks, Richardson, & Spiegelhalter, 1995), where the possibility of a future event (occurrence of a linguistic item in the case of the C-test) depends

on the previous event (linguistics item(s) preceding a gap in a C-test) and the "conditional transition probability" of one item occurring after another. This refers to the probability that a word follows a particular word, or that a word precedes a particular word or even a word before that word (see Goldsmith, 2007). With other words, predictions about the occurrence of a word can be made forward or backwards in a text based on probabilities and on the implicit statistical knowledge of the speaker, and this implicit knowledge is part of their language proficiency. Recent studies have used eye-tracking methods to investigate predictive language processing (Hopp, 2015, 2016).

Based on the discussion in the literature about the construct validity of the C-test, we argue that the underlying construct that the C-test measures is statistical knowledge about language which allows the prediction of linguistic items. This statistical knowledge is closely related to vocabulary knowledge (and crystallised intelligence), but also to linguistic processing speed (at least if the test is timed). The C-test is, therefore, not only a language test, but a test of information processing and decision-making under time pressure. We argue in the remainder of this article that this makes the C-test a good candidate for the prediction of study/training success. Various previous studies have shown that the C-test is a good predictor of academic achievement.

Daller and Xue (2009) administered a C-test to Chinese candidates (n = 23) for a place at a British university six months before they came to the UK to begin their studies. These students then undertook a one-year postgraduate programme at a business school in the UK. At the end of the year, data were collected about their study success. An operationalisation of study success in this case was the number of failed modules by the students. Some students passed all the modules at their first attempt, whereas others needed more re-sits to pass the programme. The authors assumed that those students who needed more re-sits were weaker, and that as the number of failed modules was listed in the final degree certificate, this was also an indication of their study success. A Spearman correlation between the C-test scores and the number of failed modules was negative and highly significant (rho -.565, p = .004). This means that a C-test administered more than a year before the final exams were sat, explained more than 30% of the variance in the number of failed modules. The authors also analysed essays written by the students, assessing them according to their lexical diversity and sophistication (for the use of infrequent words; in this case, words that were not in the first three frequency bands of the vocabulary programme "range", see University of Wellington, n.d.). As the C-test correlated negatively with the

number of failed modules (see above) and positively with the scores for sophistication (r = .522, $p < .05$), the authors came to the conclusion that knowledge of less frequent words is an important factor for the study success of international students, and that the C-test scores are a proxy for this knowledge.

Phelan and Daller (2013) administered a C-test to 74 international students at an induction day at the beginning of the academic year. The students were from different language backgrounds and enrolled in a variety of subjects, such as Law, Business, and Built Environment. They used the C-test scores as a predictor of the average marks that the students obtained in the first year (GPA) but could not find significant correlations at first. A second analysis revealed that some students did not attempt to sit all their modules because they were on an exchange programme and had to return to their home institution before the final exams. As a result, the authors excluded all students with a GPA below 40 to control for this issue and yielded a significant correlation between the C-test and GPA for the 44 students who sat all the exams (r = .432, $p < .01$). Interestingly, the results also revealed a high correlation between a listening task based on the IELTS format and GPA (r = .803, $p < .01$) and a high correlation between the C-test and this listening task (r = .776, $p < .001$, n = 13)[2]. As the C-test was timed and the listening task is timed per definition, other variables than vocabulary knowledge will have played a role, especially processing speed, which is related to the statistical knowledge of transitional probabilities, as mentioned in the previous section.

Daller and Yixin (2017) administered a C-test at the beginning of the academic year to 107 international students, mainly from China, from a wide range of subject areas such as English, Engineering, Mathematics, and Politics. About one-third of the sample were master's degree students, while the others were enrolled in an undergraduate programme. The aim was to predict the average marks at the end of the academic year (GPA). In addition to the C-test, the authors administered a writing task which was analysed through a series of measures of vocabulary knowledge, including Guiraud's index (see Daller, 2010). A third predictor variable was the IELTS scores of the students, obtained through a mock IELTS test. The authors carried out a series of multiple regressions and came to a final model that predicted 28.6% (R^2) of the GPA through a combination of the vocabulary measure, Guiraud's index, and the C-test scores. The IELTS scores did not make a significant further

[2] Not all students sat the listening exam

contribution to the explained variance, despite the range of scores being slightly larger than other IELTS-focused studies (range: 5.0 – 7.5, mean score: 6.14, St.Dev.: .55).

Muller and Daller (2019) used a C-test to predict training success for international trainee nurses in Australia. In total, 49 participants, mainly from China, took part in the study. Training success was operationalised in two ways: the average scores that the participants obtained in their classroom-based academic topics, and the clinical practice scores they achieved in their laboratory-based clinical topics, which also involved an assessed placement in a nursing venue, e.g. hospital placement. For the predictor variables, a C-test and IELTS test was administered at the beginning of the training year. To obtain the IELTS score, an official external test was paid for (that could be used by the candidates to help qualify for their nursing registration). The IELTS scores had similar range and greater variance to Daller and Yixin (2017) (range 5.5 – 7.5, mean score: 6.3, St.Dev.: .62). Both tests correlated significantly with the two measures of training success: IELTS academic topics (r = .509, p <.001); C-test/academic topics (r = .381, p < .01); IELTS/clinical topics (r = .302, p = .049); C-test/clinical topics (r = .417, p < .001). Interestingly, the C-test appears to be much better for the clinical topics. IELTS just achieves significance, but the C-test is a better predictor of success in the practical clinical topics. The authors of this study argue that in a nursing context, it is the combination of processing speed, general language proficiency, and in-depth conceptual knowledge (crystallised intelligence) that makes a timed C-test a good predictor of success in the clinical context, where it is necessary to "spontaneously receive and produce language in a pressured fast environment" (Muller & Daller, 2019: 7).

This can also be used as an argument for the highly predictive validity of C-tests in an academic context. International students have to process a huge amount of information in the foreign language in a short period of time to be successful as "vocabulary and ... the speed with which EFL students perform language-based tasks in English, are linked with their academic success" (Trenkic & Warmington, 2018: 13). This means that a measure that includes processing speed, such as a timed C-test, has the potential for a high predictive validity of academic or training success of international students. In order to support the arguments about the predictive validity of the C-test, we discuss in the following section two as yet unpublished projects carried out by the authors.

The hypotheses for these two studies are based on the literature discussed so far.

Hypotheses

1. C-test scores at the beginning of an academic year correlate highly with the marks obtained at the end of the year (both studies)
2. This also holds for a variety of linguistic and literature topics and test formats (study 2)

Study 1

We repeated the study by Daller and Yixin (2017) (see literature review above) with the academic cohort of the following year.

Participants

For the repeat study, we had 134 participants in total at a British university (95 Chinese, 16 other Asian, 8 African, and 15 European students). The average age was 22.93 (St.Dev. 4.54), and around one-third were undergraduate students and two-thirds postgraduates, from a range of disciplines such as Media, Engineering, Mathematics, Politics, and Linguistics.

Measures and Procedure

The same C-test as in Daller and Yixin (2017) was used with all participants both at the beginning of the academic year and in May towards the end of the academic year shortly before the students sat the exams in their different disciplines. We obtained the marks that the students received after the first semester, the final mark at the end of the academic year, and the number of failed modules in June at the end of the academic year. The C-test was administered under the supervision of one of the researchers, and the students were given 30 minutes for the five sub-texts.

Results

Table 1 shows the correlations between the two C-test scores, the mark at the end of the first semester (Marks Sem 1), the final mark, and the number of failed modules (Pearson correlations). Note that the sample was smaller than 134 in some cases because participation was voluntary, and some students did not sit the second test round.

Table 1

Correlations between C-tests and marks / failed modules

	C-test Sept	C-test May	Marks Sem 1	Final Marks	# Failed modules
C-test Sept	-	.779** n = 48	.420** n = 57	.451 ** n = 134	-.289** n = 134
C-test May		-		.534 ** n = 48	-.408** n = 48
Marks Sem 1			-	.758** n = 56	-.530** n = 56
Final Marks				-	-.623** n = 134
# Failed modules					-

** $p < .01$

The first point of note is that there is a strong correlation ($r = .758$) between the marks obtained after the first semester and the final marks. Not surprisingly, students who did well in the first semester also did well in their final marks. Both C-test scores correlated significantly with the marks in both semesters. The C-test administered towards the end of the academic year correlated highly with the final marks ($r = .534$, $p < .001$). What is more surprising is that the C-test administered at the beginning of the academic year (C-test September) also correlated significantly with the final marks and explained around 20% of the variance of the final marks ($r = .451$, $p < .001$). Both C-tests predicted the number of

failed modules, and the C-test May explained approximately 16% of the variance in the number of failed modules at the end of the academic year.

Conclusions for Study 1

Study 1 supports the findings of Daller and Yixin (2017). A C-test administered at the beginning of an academic year is a good predictor of the final marks, even in a wide range of subject areas. This means there must be a construct that underlies different exam types and subject areas. This could have consequences for the admissions process in HE and elsewhere, and for identifying students at risk. Although most participants in Study 1 were from a Chinese background, these findings should be generalisable to other L1 speakers.

Study 2

Context

This study (Daller, Vaatstra, & Verspoor, in preparation) is different from previous studies as most speakers had an L1 that was close to English or were even native English speakers. The students followed a degree programme that consisted of courses in three different disciplines: modern English literature, medieval English literature, and English linguistics. In addition, the students took courses in English for Academic Purposes (EAP) to support their writing and speaking skills in an academic context.

Participants

The participants in this study consisted of 89 first-year students enrolled in the Bachelor's degree programme in English Language and Culture at a university in the Netherlands. Overall, 80% of the sample consisted of L1 speakers of Dutch. The remaining 20% were international students, one-third of whom were native speakers of English, one-third L1 speakers of German, and the final group had another language as their L1.

Measures and Procedure

A C-test was administered in the introduction week prior to the first semester. The final examinations were scheduled at the end of the teaching weeks for each course, and so took place in 10-week intervals throughout the academic year.

The academic year was made up of several introductory courses to each field of study, adding up to 12 courses in total. The final grade for each course was generally a weighted average of several components, as most courses used continuous testing in the teaching weeks, followed by a final examination at the end of the teaching period. Most exams made use of multiple-choice quizzes and exams, which was complemented by short papers in the second semester. The modern and medieval literature courses generally required their students to hand in a written assignment during the teaching weeks, and always ended with a written exam consisting of open questions. The English for Academic Purposes (EAP) programme focused on writing and speaking skills, so students wrote multiple assignments throughout each course, with a revised text or written exam as their final examination. With the exception of the first teaching block, the students were also graded on their oral skills at the end of each EAP course.

Results

It was found that the C-test was highly reliable (Cronbach's alpha = .833, 5 items), with the exclusion of any sub-test decreasing the Cronbach's alpha value. We therefore took all five sub-tests together in the following computations. Figure 1 shows the spread of the C-test scores in the five sub-items.

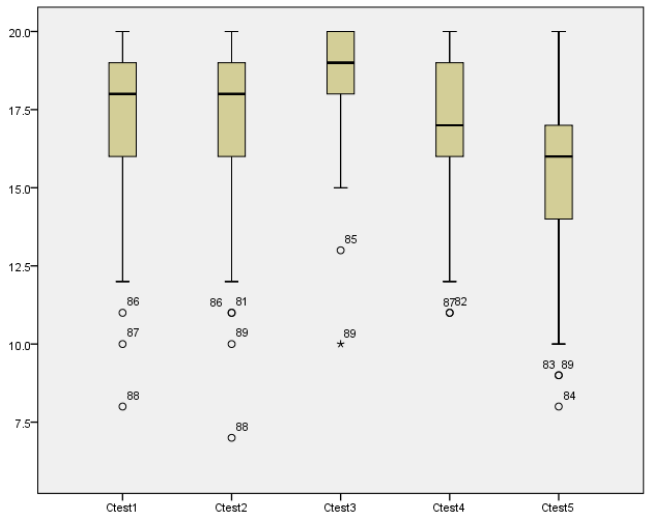

Figure 1
Spread of the C-test scores

Figure 1 shows that there is potentially a ceiling effect, which is not surprising given that the first languages of the participants were closely related to English or even English as a first language. Nevertheless, the C-test was found to be highly reliable which showed that it is a robust test even for this group of participants.

To investigate the predictive validity of the C-test, we added all the 12 exam marks together to achieve an overall mark for the participants as an indicator of their study success. The correlation between the total C-test scores and the overall marks was highly significant ($r = .358$, $p < .01$, $n = 55$). Note that the overall size of the sample was smaller than the number of participants (89) because there were some missing values in the 12 exams.

Discussion of Study 2

The findings from Study 2 are important for two reasons. Firstly, they show that the C-test is a robust test format even under conditions that normally have negative effects on the reliability of a test (ceiling effects). The study also shows that the C-test has a high predictive validity even for a large variety of exam settings, topics, and scoring procedures. This is an indication that the C-test taps into a general underlying proficiency for these exams.

Overall Discussion

Based on the literature review, the authors came to the hypothesis that the C-test is a good predictor of study success. This is supported by the two studies reported above. Whether it is the study success of international, mainly Chinese, students (Study 1) or native/near native speakers (Study 2), the C-test is a good predictor in a large variety of subject areas. This raises questions about the construct that underlies the successful predictions in these contexts. First of all, we argue that it is in-depth vocabulary knowledge in context which is related to crystallised intelligence. However, it is not enough to have this knowledge, and a further aspect related to vocabulary knowledge is important when a test has to be completed under time pressure. It can be safely assumed that linguistic processing speed and vocabulary knowledge are related, and that a person who knows more words in context has a higher linguistic processing speed. This processing speed is based on the statistical knowledge of conditional transitional probabilities between linguistic items. It is part of a high language proficiency to be able to predict which item could follow after a single word or string of words. Earlier in the literature, this has been called "expectancy grammar" (Oller, 1976), although operationalisation of the term was missing. These two aspects of language proficiency, vocabulary knowledge and linguistic processing speed, together are necessary to fill in a C-test and to be successful in a study context. To be a successful student, one has to process information in a short period of time, e.g. understanding a lecture or a course book. Trenkic and Warmington (2018) have shown that many international students have a slower linguistic processing speed in the foreign language which explains why many of them struggle at university level. However, the C-test also seems to measure processing speed in general, as argued by Muller and Daller (2019), although it is difficult to disentangle general from linguistic processing speed. Muller and Daller (2019) demonstrate that C-test scores predict success in an environment where the ability to process information rapidly, namely spontaneous face-to-face communication in a clinical context, is important.

Earlier studies have argued that the C-test measures "general language proficiency", as C-test scores correlate with test scores in the four classical language skills (reading, writing, listening, and speaking). However, this assumption of a unified underlying concept has not been explored in detail. The findings of the different studies discussed in this article have shed more light on this underlying common concept. Future research needs to investigate whether linguistic processing speed and vocabulary knowledge in context can be teased apart for non-native speakers. Wockenfuß and Raatz (2014) showed this for native

speakers, and a similar design would be necessary to investigate this for non-native speakers. In line with Trenkic and Warmington (2018), one could expect that the potentially high non-verbal processing speed of non-native speaking students is not reflected in the linguistic processing speed in the foreign language (due to smaller vocabulary sizes). The question of whether there are one or more underlying constructs for the C-test cannot be answered at the moment, since all the factors are intertwined, but at least the candidates for a definition of this construct can be named: statistical knowledge, vocabulary knowledge in context, and linguistic processing speed. Further eye-tracking studies are a promising way to further investigate the processes that take place when C-tests are taken. Additional information about crystallised intelligence and (non-verbal) processing speed might be useful in these studies to draw a more fine-grained picture of the construct validity of the C-test format.

Conclusions

The discussion of the studies in this article make it clear that the C-test format can be used effectively to predict the study success of international students, be it in an academic or a training context. This has pedagogical consequences as candidates who need additional language support can be identified before they begin their studies with a quick and easy test. It could also be used to complement admissions tests when time and resources are limited. Whether it can replace existing admissions procedures is a question for future research. The high correlation of C-tests with test scores in the four classical skills (Eckes & Grotjahn, 2006; Klein-Braley, 1985) point in this direction. As an admissions test, it would need to be administered under controlled conditions, including a time limit. One limitation might be the low face validity of the C-test. For almost all participants in the reported studies, the C-test was an unknown test format. However, if it was to be more widely used, including as part of training programmes in English, this limitation could be overcome.

References

Alderson, C. (2002). Testing proficiency and achievement: principles and practice. In J. Coleman, R. Grotjahn, & U. Raatz (Eds.), *University language testing and the C-test* (pp. 15–30). Bochum: AKS-Verlag.

Babaii, E., & Ansary, H. (2001). The C-test: a valid operationalization of reduced redundancy principle? *System*, *29*(2), 209-219.

Baghaei, P., & Tabatabaee, M. (2015). The C-test: an integrative measure of crystallized intelligence. *Journal of Intelligence*, 3, 46-48.

Bolten, J. (1992). Wie schwierig ist ein C-Test? Erfahrungen mit dem C-Test als Einstufungstest in Hochschulkursen Deutsch als Fremdsprache. In R. Grotjahn (Eds.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen, 2*, 193-203.

Boyd, J. K., & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, *87*(1), 55 – 83.

Daller, M. H., & Xue, H. (2009). Vocabulary knowledge and academic success: A study of Chinese students in UK higher education. In B. Richards, H. Daller, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition* (pp. 179-193). London: Palgrave Macmillan.

Daller, M. (2010). *Guiraud's index*. Retrieved from http://eprints.uwe.ac.uk/id/eprint/11902

Daller, M. H., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review, 4*(1), 173-193.

Daller, M., & Yixin, W. (2017). Predicting study success of international students. *Applied Linguistics Review*, *8*(4), 355-374.

Daller, Vaatstra & Verspoor (in preparation). The predictive validity of the C-test for students' study success.

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*(2), 187-206.

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290-325.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (Eds.). (1995). *Markov chain Monte Carlo in practice*. Place?: CRC Press.

Goldsmith, J. (2007). Probability for linguists. *Mathématiques et sciences humaines (Mathematics and social sciences)*, (180), 73-98.

Grotjahn, R. & Allner, B. (1996). Der C-Test in der Sprachlichen Aufnahmeprüfung an Studienkollegs für ausländische Studierende an Universitäten in NordrheinWestfalen. In R.

Grotjahn (Eds.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3, pp. 279–335). Bochum: Brockmeyer.

Hastings, A. J. (2002). In defense of C-testing. In R. Grotjahn (Eds.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 11–29). Bochum: AKS-Verlag.

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, *36*(2), 129-173.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277-307.

Huhta, A. (1996). Validating an EFL C-test for students of English philology. In R. Grotjahn (Eds.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3, pp. 197–229). Bochum: Brockmeyer.

Jafarpur, A. (2002). A comparative study of a C-test and a cloze test. In R. Grotjahn (Eds.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 31–51). Bochum: AKS-Verlag

Klein-Braley, C. (1985). A cloze-up on the C-Test: a study in the construct validation of authentic tests. *Language Testing*, *2*(1), 76-104.

Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: An appraisal. *Language Testing*, *14*(1), 47-84.

Muller, A. & Daller, M. (2019). Predicting international students' clinical and academic grades using two language tests (IELTS and C-test): a correlational research study. *Nurse Education Today* 72, 6-11.

Oller, J. W. (1976). Evidence for a general language proficiency factor: An expectancy grammar. *Die neueren sprachen*, *75*(2), 165-174.

Raatz, U., & Klein-Braley, C. (1981). The C-Test--A Modification of the Cloze Procedure. ERIC.

Singleton, D., & Singleton, E. (2002). The C-test and L2 acquisition/processing research. In J. Coleman, R. Grotjahn, T. & U. Raatz (Eds.), *University Language Testing and the C-test* (pp. 143-168). Bochum: AKS-Verlag.

Spolsky, B. (1985). What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing*, *2*(2), 180-191.

TakeIELTS (2019). Retrieved January 9, 2020, from https://takeielts.britishcouncil.org/about/press/ielts-grows-three-half-million-year

Trenkic, D., & Warmington, M. (2018). Language and literacy skills of home and international university students: How different are they, and does it matter?. *Bilingualism: Language and Cognition*, 1-17.

UNESCO (2019). Retrieved January 9, 2020, from http://data.uis.unesco.org/Index.aspx?queryid=172

Universities UK, International (2019). Retrieved January 9, 2020, from https://www.universitiesuk.ac.uk/International/Documents/2019/International%20facts%20and%20figures%20slides.pdf

University of Wellington (n.d.). Retrieved January 9, 2020, from https://www.wgtn.ac.nz/lals/resources/range

Wang-Taylor, Y., & Milton, J. (2019). Predicting international students' academic success with vocabulary and intercultural communicative competence. *Journal of Applied Linguistics*, (32), 83-101.doi: https://ejournals.lib.auth.gr/jal/article/view/7533

Wockenfuß, V., & Raatz, U. (2014). Zur Validität von muttersprachli-chen C-Tests: Bedeutung von verbaler Intelligenz und Informationsverarbeitungsgeschwindigkeit unter Berücksichtigung des Lebensalters. In Grotjahn, R. (Ed.) *Der C-Test: Aktuelle Tendenzen/The C-Test: Current Trends*, 191-224.

Yixin, W., & Daller, M. (2015). Predicting Chinese Students' Academic Achievement in the UK. *Learning, Working and Communicating in a Global Context, 217*.